

Claims:

1. An information search system for crawling a web site via a network, comprising:

structure analyzing means for analyzing a structure of source code in a prescribed web page;

significance calculating means for calculating a degree of significance of a web site linking from said prescribed web page, based on an analysis result of said structure analyzing means; and

crawling means for crawling the web site depending on the degree of significance calculated by said significance calculating means.

2. An information search system according to Claim 1, wherein said structure analyzing means associates mutually relevant information elements with each other, among information elements contained in said source code.

3. An information search system according to Claim 1, wherein said significance calculating means calculates the degree of significance of said web site selectively using a strategy that is for calculating the degree of significance

of said web site, from among strategies that are provided in advance.

4. An information search system according to Claim 3, wherein said significance calculating means selects plural strategies as strategies for calculating the degree of significance of said web site, and uses them by giving weights thereto, respectively.

5. An information search system comprising:
document structure analyzing means for analyzing a document structure of an HTML document, and adding an information element acquired by the analysis to a corresponding anchor; and

crawling means for crawling a web site linking from said anchor, depending on a degree of significance of said anchor calculated based on said information element acquired through the analysis of said document structure analyzing means.

6. An information search system according to Claim 5, wherein said document structure analyzing means groups

respective information elements forming said HTML document into blocks each unified in terms of a meaning of said information elements, and adds the information element in each block to an anchor in the same block as additional information.

7. An information search system according to Claim 5, further comprising significance calculating means for calculating a degree of significance of said anchor based on said information element acquired through the analysis of said document structure analyzing means and according to a preselected prescribed strategy,

wherein said crawling means crawls the web site depending on the degree of significance of said anchor calculated by said significance calculating means.

8. An information search method for crawling a web site via a network using a computer, said method comprising the steps of:

acquiring a web page as initial information and storing source code into a storage device;

reading the source code of said web page from said

storage device, conducting a structure analysis of said web page, and storing a result of the analysis into said storing device;

calculating a degree of significance of a web site linking from said web page, based on the result of said structure analysis stored in said storage device; and

accessing the web site depending on the calculated degree of significance to acquire contents thereof, and storing them into said storage device.

9. An HTML document structure analyzing method using a computer, said method comprising the steps of:

reading an HTML document being a processing object from a memory, blocking information elements forming said HTML document based on tags of said HTML document, and storing blocked structural data of said HTML document into the memory; and

reading the blocked structural data of said HTML document from said memory, updating block structures of said HTML document by associating the information elements that are mutually relevant in terms of a meaning, and storing the updated structural data into the memory.

10. A program product for controlling a computer connected to a network so as to crawl a web site, said program product causing said computer to execute:

a process of acquiring a web page as initial information and storing source code into a storage device;

a process of reading the source code of said web page from said storage device, conducting a structure analysis of said web page, and storing a result of the analysis into said storing device;

a process of calculating a degree of significance of a web site linking from said web page, based on the result of said structure analysis stored in said storage device; and

a process of accessing the web site depending on the calculated degree of significance to acquire contents thereof, and storing them into said storage device.

11. A program product according to Claim 10, wherein said program product causes said computer to conduct said structure analysis by associating mutually relevant information elements with each other, among information elements contained in said source code.

12. A program product according to Claim 10, wherein, in the process of calculating the degree of significance of said web site, plural strategies are used as strategies for calculating the degree of significance of said web site, by giving weights thereto, respectively.

13. A program product for controlling a computer so as to analyze an HTML document structure, said program product causing said computer to execute:

a first process of reading an HTML document being a processing object from a memory, blocking information elements forming said HTML document based on tags of said HTML document, and storing blocked structural data of said HTML document into the memory; and

a second process of reading the blocked structural data of said HTML document from said memory, updating block structures of said HTML document by associating the information elements that are mutually relevant in terms of a meaning, and storing the updated structural data into the memory.

14. A program product according to Claim 13, wherein, in said second process by said program product, said program product causes said computer to execute:

a process of identifying an unnecessary information element in terms of a purpose of a document structure analysis;

a process of deleting a block having no structural meaning;

a process of merging said information elements or dividing a block based on contents of said information elements; and

a process of merging the block structures based on information contained in each block.